

# **Intervalos de predicción de la frecuencia de accidentes en tramos de la Red de Carreteras Española, para flujos heterogéneos. Aplicación a ejes de transporte de mercancías**

**Blanca Arenas Ramírez**

Directora de la Unidad de Estudios de Transportes (INSIA), ETSII. UPM,

.e-mail: [barenas@etsii.upm.es](mailto:barenas@etsii.upm.es)

Teléfono: 91 - 3363014. Fax: 91 - 3365302

**Francisco Aparicio Izquierdo**

Catedrático de la Escuela Técnica Superior de Ingenieros Industriales de Madrid (UPM) y Director del Instituto Universitario de Investigación del Automóvil (INSIA). Madrid.

España

e-mail: [francisco.aparicio@upm.es](mailto:francisco.aparicio@upm.es) .

Teléfono: 91 – 3365303. Fax: 91 - 3365302

**Camino González Fernández**

Profesora Titular de Estadística de la ETSII. UPM. Profesora Titular de Estadística de la ETSII. UPM. Madrid. España.

e-mail: [camino@etsii.upm.es](mailto:camino@etsii.upm.es)

Teléfono: 91 - 3363149.

## **RESUMEN**

Los modelos lineales generales, tanto en sus alternativas de Poisson como de binomial negativa, son considerados los más adecuados por los analistas para relacionar la frecuencia o tasa de accidentes de tráfico con diferentes variables explicativas (flujos de tráfico, geometría, tipos de vía, etc.) y en diferentes escenarios.

En este trabajo se desarrollan los intervalos de confianza y de predicción basados en un modelo con flujo de tráfico heterogéneo, en tramos de la Red de Carreteras Españolas, y se aplican al eje de transporte de mercancías Madrid-Barcelona (MB), muy importante, tanto por el volumen de tráfico como por la distancia media?

La construcción de estos intervalos permite la predicción de la frecuencia o tasa de accidentes para tramos caracterizados por diferentes valores de las variables explicativas, permitiendo generalizar el modelo y asignar intervalos de confianza a cada tramo.

Se han formulado tres tipos de intervalos de predicción para:

- El número medio de accidentes o *seguridad global* en tramos con determinadas características.
- El número medio de accidentes o *seguridad específica* en un tramo diferente con determinadas características. Este intervalo solo es posible calcularlo si se considera un modelo de binomial negativa.
- El número de accidentes en un tramo diferente con determinadas características.

La aplicación a tramos concretos como es el caso de los tramos del eje de mercancías Madrid-Barcelona, permite validar el modelo desarrollado utilizando los datos de diferentes años; permite evaluar el comportamiento de la accidentalidad en el eje

Madrid-Barcelona, y analizar como ésta se modifica si se plantean diferentes escenarios de flujos heterogéneos.

## 1. PRESENTACIÓN

Los modelos lineales generales de Poisson y binomial negativa, tienen gran aceptación entre los analistas, para determinar la relación entre la tasa o frecuencia de accidentes de tráfico y distintas variables de influencia como son composición y volumen de tráfico o la geometría de la vía.

En este trabajo se muestran los resultados del ajuste de modelos de la binomial negativa para relacionar la frecuencia de accidentes totales, con variables de tráfico y de tipo de vía en tramos de distintas categorías de vías de la Red de Carreteras del Estado (RCE), y de la aplicación de los intervalos teóricos de predicción de los distintos niveles de seguridad que se pueden definir en un modelo de binomial negativa. Se han utilizado para el ajuste de modelos los datos de accidentes y de tráfico correspondientes al año 2001.

De entre los modelos de binomial negativa ajustados, que se diferencian en las variables explicativas incorporadas, se ha seleccionado el mejor, en base a criterios estadísticos, siguiendo algunos de los recopilados en Hardin y Hilbe (2007) (AIC, BIC, soporte, desviación, etc.) y considerando su adecuación a la realidad que quiere reproducir.

El modelo seleccionado se ha validado utilizando los datos correspondientes a los años 2002 y 2003. Posteriormente se ha aplicado para la predicción de la tasa y del número de accidentes, en distintos escenarios hipotéticos contruídos con distintos valores de los regresores, en tramos genéricos de vías interurbanas de autopistas (AP), de autovías (AV), vías de doble calzada (DC) y vía convencional (C). Los intervalos se ha establecido para la seguridad global o media en tramos con determinadas características; la seguridad específica de un tramo, y el número de accidentes totales en un tramo específico. (McGullagh y Nelder, 1989, Hauer y otros, 1988,, Wood G. R., 2005, Myers R. y Montgomery, D., 1997).

Finalmente, la metodología descrita se ha aplicado para la determinación de los intervalos de confianza y predicción de la tasa y número de accidentes totales, en los tramos del eje de transporte Madrid- Barcelona (MB).

## 2. METODOLOGÍA Y DATOS

La expresión más general para predecir el número de accidentes es:

$$\hat{E}[Y] = \exp(\hat{\beta}_0 + \hat{\beta}_1 \log(AADT) + \hat{\beta}_2 \log(\%HGV) + \hat{\alpha}_i \text{via}_i + \hat{\delta}_i \log(\%HGV)_i + \log(vk))$$

$i = AP, AV, DC$

donde AADT es la intensidad media diaria anual de vehículos medida en cada tramo aforado, %HGV es el porcentaje de vehículos industriales de transporte de mercancías determinado como  $100 * AADTHGV / AADT$ , con AADTHGV la intensidad media diaria anual de vehículos industriales, vk es la medida de exposición adoptada en millón de vehículo-kilómetro y determinada como  $365 * AADT * \text{longitud del tramo}$ , y vía la clasificación funcional del tipo de vía, según la denominación adoptada en el Mapa de

Tráfico del año 2001, del cual se ha obtenido la base de datos de tramos y los tráficos medidos.

El modelo incorpora términos de interacción entre los diferentes regresores por ejemplo el término del  $\log(\%HGV)$  indica que la influencia del % de vehículos industriales sobre el número de accidentes será diferente dependiendo del tipo de vía que se esté analizando.

La ecuación general del modelo, estimada utilizando el modelo de binomial negativa, se puede particularizar por tipo de vía y expresar en términos de la tasa de accidentes o número medio de accidentes por unidad de exposición- calculada como  $\hat{\lambda}_{X_j} = \frac{E(X_j)}{v k_j}$ .

Las expresiones de esta tasa por tipo de vía son:

$$\begin{aligned}\hat{\lambda}_{U_j} &= 0.2736 \cdot (AADT_j)^{-0.000001} \cdot (HGV_j)^{0.0638} \\ \hat{\lambda}_{AV_j} &= 0.4368 \cdot (AADT_j)^{-0.000001} \cdot (HGV_j)^{0.0103} \\ \hat{\lambda}_{X_j} &= 0.3405 \cdot (AADT_j)^{-0.000001} \cdot (HGV_j)^{0.2491} \\ \hat{\lambda}_3 &= 1.2402 \cdot (AADT_j)^{-0.000001} \cdot (\%HGV_j)^{0.1639}\end{aligned}$$

A partir de la teoría de los modelos lineales generales, se pueden desarrollar los diferentes intervalos de predicción. En la Tabla 1 se muestran los límites de los intervalos de predicción del 95% para el número medio de accidentes por unidad de exposición o tasa global de accidentes  $\hat{\lambda}_n$ , para el número medio de accidentes dada una unidad de exposición (la de cada tramo) o tasa específica de cada tipo de tramo  $\Lambda_h$ , y para el número de accidentes dada una unidad de exposición en un tramo con determinadas características  $Y_h$ .

Parámetro	Límites Intervalos
$\hat{\lambda}_n$	$\left[ \frac{\hat{\lambda}_h}{e^{1.96\sqrt{\text{var}(\hat{\eta})}}}, \hat{\lambda}_h \cdot e^{1.96\sqrt{\text{var}(\hat{\eta})}} \right]$
$\Lambda_h$	$\left[ \max \left\{ 0, \hat{\lambda}_h - 1.96\sqrt{\hat{\lambda}_h^2 \text{var}(\hat{\eta})} \right\}, \hat{\lambda}_h + 1.96\sqrt{\hat{\lambda}_h^2 \text{var}(\hat{\eta})} \right]; \left[ \frac{\hat{\lambda}_h^2 \text{var}(\hat{\eta}) + \hat{\lambda}_h^2}{\hat{\lambda}_h}, \frac{\hat{\lambda}_h^2 \text{var}(\hat{\eta}) + \hat{\lambda}_h^2}{\hat{\lambda}_h} \right]$
$Y_h$	$\left[ 0; v k \cdot \hat{\lambda}_h + 1.96\sqrt{v k \cdot \left( \hat{\lambda}_h + \hat{\lambda}_h^2 \text{var}(\hat{\eta}) \right)} \right]; \left[ \frac{\hat{\lambda}_h^2 \text{var}(\hat{\eta}) + \hat{\lambda}_h^2}{\hat{\lambda}_h}, \frac{\hat{\lambda}_h^2 \text{var}(\hat{\eta}) + \hat{\lambda}_h^2}{\hat{\lambda}_h} \right]$

**Tabla 1 – Intervalos de predicción del 95% del modelo de la binomial negativa.**

**Nota:**  $\text{var}(\hat{\eta}) = x_0 (X'VX)^{-1} x_0$ , entre corchetes se indican los valores posibles del que se elige el valor mayor de los dos.

### 3. RESULTADOS

Las predicciones del número de accidentes  $\hat{Y}_h$  así como los límites del intervalo de predicción de la estimación puntual, del Eje Madrid Barcelona para los años 2001-2003 se muestran en las figuras 1, 2 y 3. Estas figuras se presentan para los distintos tramos que componen el eje, organizados por orden creciente de la predicción del número de accidentes, es decir, en primer lugar se muestran los tramos de **AP**, seguidos por los de autovía **AV** y el gráfico cierra con las predicciones en tramos de vía convencional **C**.

Los valores de accidentes observados  $y_h$  durante estos años, en los tramos que componen el eje están incluidos en el intervalo de predicción proporcionado por el modelo, con la excepción de un número muy reducido de tramos ( cinco en el año 2001, seis en el año 2002 y cinco en el 2003, de un total de 81), en los cuales los valores observados no quedan dentro de los límites de la predicción.

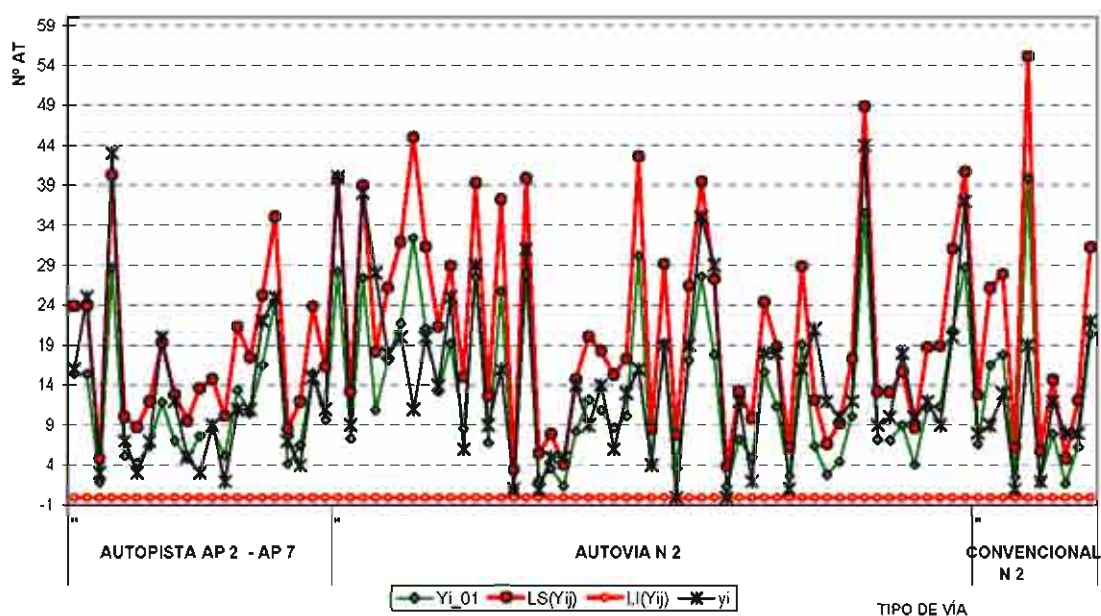
En la figura 4 se muestra el número total de accidentes pronosticado y observados de los años 2001, 2002 y 2003 del eje Madrid- Barcelona, y por tipo de vía. Globalmente, el número de accidentes totales pronosticados por el modelo difiere con los observados en el 10% en los años 2001 y 2002, y la diferencia en el año 2003, es del 2%.

Por tipo de vía, el modelo arroja resultados mejores en autopistas y autovías, que en vías convencionales. En este último tipo de vías el error en la estimación puntual es del 20% para los años 2001 y 2003, mientras que el error para el año 2002 fué del 37%.

Detrás de las mayores diferencias en el año 2002, pueden puntualizarse dos factores, no poco importantes: por un lado los datos para algunos tramos del eje, de este tipo de vías son incompletos, al no haber sido posible obtener los datos del recuento de vehículos del Mapa de tráfico en la sección transversal correspondiente y, por tanto, los valores han sido interpolados. Estos datos se introducen directamente en el modelo a través de las variables AADT y %HGV.

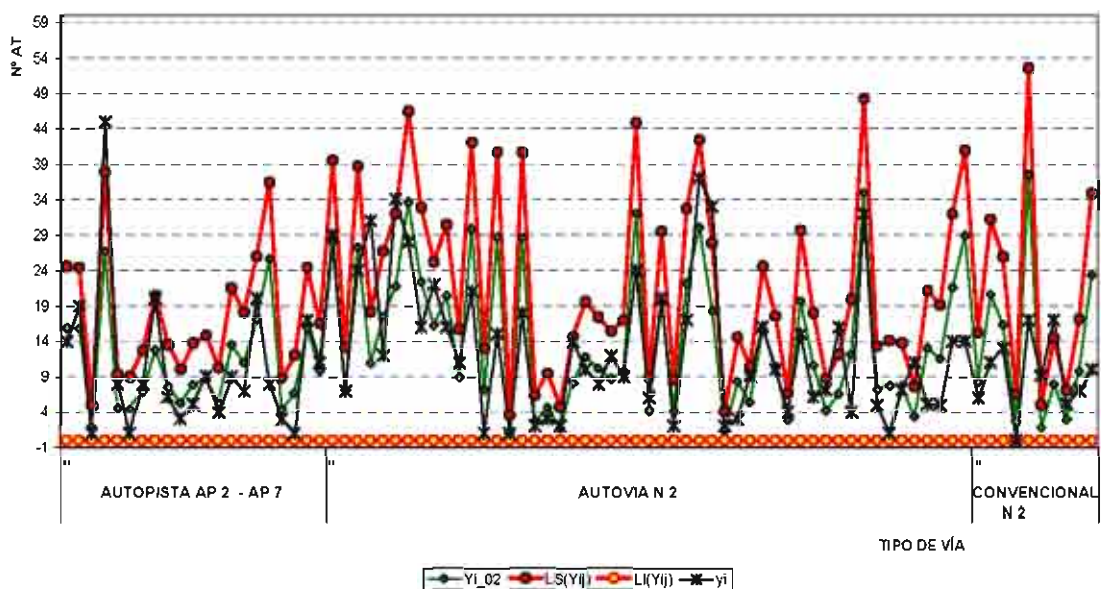
Por otro lado, la falta de información de tráfico del tramo no permitió, por tanto, la determinación del número de accidentes en ellos, mediante la aplicación del algoritmo de búsqueda y recuento de accidentes, y los valores de accidentes han tenido que ser interpolados entre los del año anterior y posterior. Las diferencias por tanto, deben ser consideradas con precaución.

PREDICCIÓN DEL NÚMERO DE ACCIDENTES TOTALES EN LOS TRAMOS DEL EJE MADRID-BARCELONA. OBSERVADOS 2001



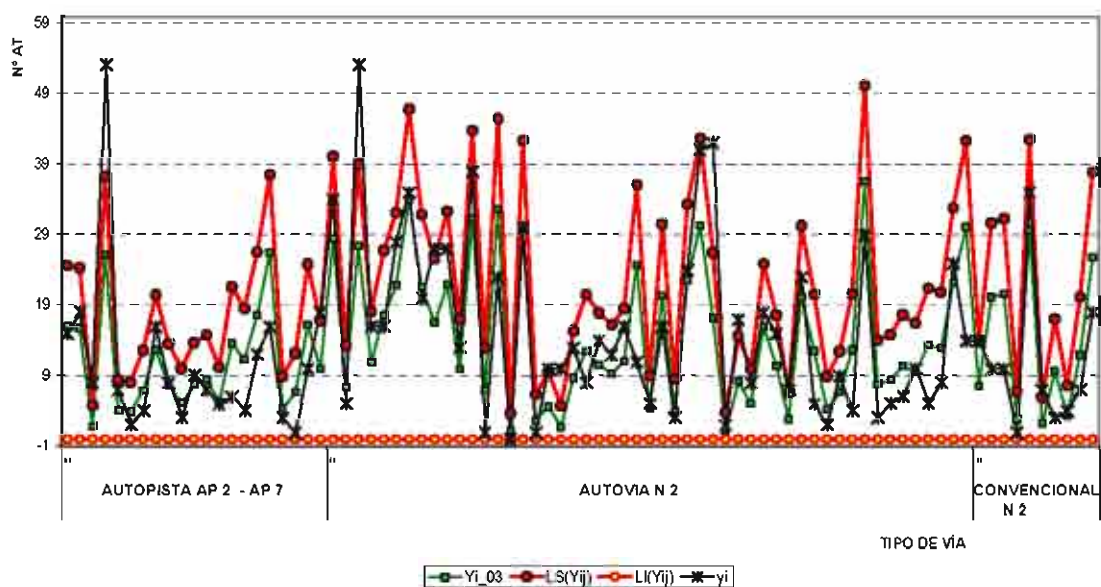
**Figura 1 – Predicción del número de accidentes en los tramos del Eje MB. Límites del intervalo de predicción (95%). Observados. Año 2001.**

PREDICCIÓN DEL NÚMERO DE ACCIDENTES TOTALES EN LOS TRAMOS DEL EJE MADRID-BARCELONA- OBSERVADOS 2002.



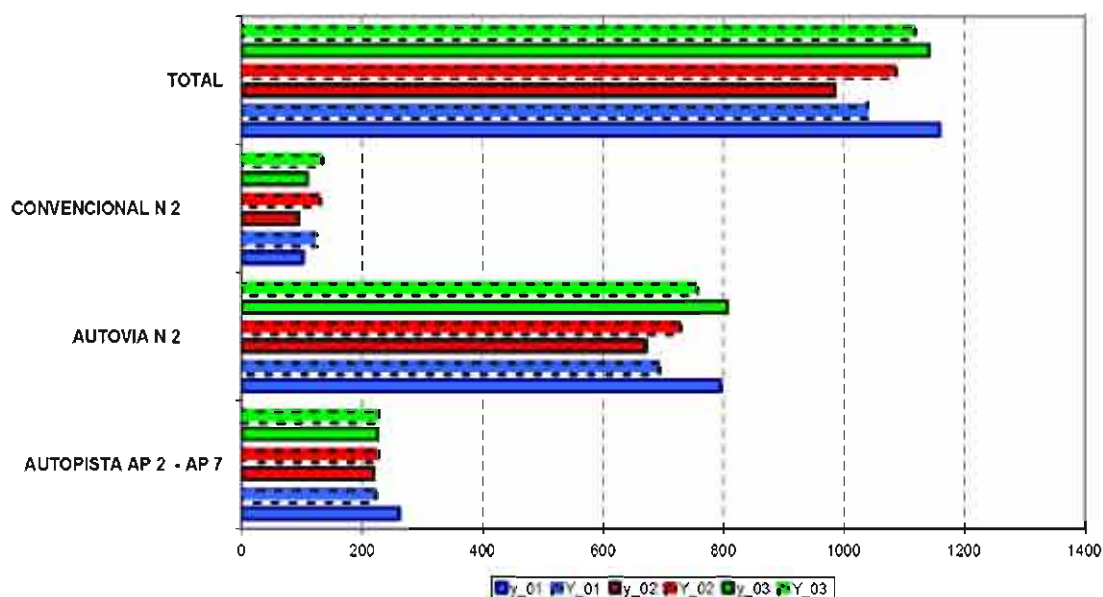
**Figura 2 – Predicción del número de accidentes en los tramos del Eje MB. Límites del intervalo de predicción (95%). Observados. Año 2002**

PREDICCIÓN DEL NÚMERO DE ACCIDENTES TOTALES EN LOS TRAMOS DEL EJE MADRID-BARCELONA- OBSERVADOS 2003.



**Figura 3 – Predicción del número de accidentes en los tramos del Eje MB. Límites del intervalo de predicción (95%). Observados. Año 2003**

Número total de accidentes en el eje MB. Predicción ( $Y_i$ ) frente a Observados ( $y_i$ )  
Total y por tipo de vía. Año 2001– 2003



**Figura 4 – Número total de accidentes en el eje MB. Predicción  $Y_i$  frente a Observados  $y_i$ . Total y por tipo de vía. Año 2001-2003.**



#### 4. DISCUSIÓN Y CONCLUSIONES

En este trabajo se presenta la aplicación de modelos lineales generales en su variante de binomial negativa para modelar y predecir el número de accidentes en diferentes tramos en función de volumen de tráfico, composición de tráfico y tipo de vía. Se ha ajustado un modelo utilizando los datos correspondientes a 2001.

La predicción puntual proporcionada por el modelo se ha completado incorporando la incertidumbre asociada a la fenomenología del propio suceso “accidente” y la asociada a la estimación de los parámetros del modelo dando lugar a las expresiones de los diferentes intervalos de predicción. No solo se dispone por tanto de una estimación puntual del número de accidentes sino de un intervalo para dicha variable.

Las predicciones e intervalos del número de accidentes calculados mediante la metodología que aquí se presenta permite establecer la validez de la aplicación del modelo ajustado a nuevos datos y nuevos valores de las variables independientes.

Como ilustración se ha aplicado el modelo ajustado con datos del año 2001 a la predicción de accidentes en los tramos del eje de transporte de mercancías Madrid Barcelona, con nuevos datos de tráfico, para los años 2002 y 2003. La aplicación del modelo es satisfactoria en tanto que los valores pronosticados por el modelo, no ofrecen diferencias sustanciales con los observados en dicho período, y los valores observados  $y_h$  durante estos años, en los tramos que componen el eje están incluidos en el intervalo de predicción proporcionado por el modelo, con la excepción de un número muy reducido de tramos, en los cuales los valores observados no quedan dentro de los límites de la predicción. Globalmente, los totales pronosticados por el modelo difieren con los observados en el 10% en los años 2001 y 2002, y la diferencia en el año 2003, es del 2%. Por tipo de vía, el modelo arroja resultados mejores en autopista y autovía, que en vía convencional.

La determinación de los diferentes intervalos de predicción para el número medio de accidentes y para el número de accidentes en un nuevo tramo con unas determinadas características, permite utilizar la metodología desarrollada (modelo e intervalos de predicción) como herramienta para la gestión de políticas de seguridad vial en base al comportamiento diferenciado de la accidentalidad por tipo de vía, ya que incorpora la incertidumbre asociada a la variabilidad del tráfico y de la exposición.

La incorporación de nuevas variables explicativas al modelo de predicción ajustado podría contribuir a mejorar los resultados.

#### 5. REFERENCIAS

McCULLAGH P., NELDER J.A. (1989). *Generalized Linear Models*. 2nd Ed. Chapman&Hall/CRC, Florida.

HAUER E., NG J.C.N., LOVELL J. (1988). Estimation of safety at signalized intersections. Transportation Res. Rec. 1185. 48-61.

WOOD G. R. (2005). Confidence and Prediction Intervals for generalised linear accident models. *Accident Analysis and Prevention* 37. 267-273.

MYERS R. , MONTGOMERY D. (1997). A tutorial on Generalized Linear Models. *Journal of Quality Technology*. 29(3). 274-291.

HARDIN, J.W. y HILBE, J.M. (2007). *Generalized Linear Models and Extensions* Stata Press.